J. Appl. Cryst. (1988). **21**, 925–928

Versatile Rigid-Body and Group Refinement in Real Space

BY T. O. YEATES* AND D. C. REES

UCLA Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, CA 90024, USA

(*Received* 20 *January* 1988; *accepted* 22 *June* 1988)

Abstract

A procedure is described for performing least-squares rigid-body or group refinements in real space. The necessary derivatives are obtained from gradients in a difference Fourier map [Jack & Levitt (1978). *Acta Cryst.* **A34**, 931–935] allowing the use of fast Fourier transforms for rapid calculations. The computation time is nearly independent of the number of rigid groups being refined. The orientational transformation of the molecule or group is represented in a linear matrix form. The inherent non-linearity is treated outside of the refinement calculations, and trigonometric calculations are avoided. The linear approximation provides versatility in the types of constraints that may be applied during the refinement. Errors in unit-cell parameters may be detected through the strain component of the transformation matrix. The results of refinement of two protein structures are described as examples.

1. Introduction

Following approximate placement of a molecular replacement model, one frequently optimizes the position and orientation of the model in the unit cell. This step is important because initial minimization of the discrepancy between observed and calculated structure factors as a function of a small number of variables is expected to improve the radius of convergence for the atomic refinement that may follow (Scheringer, 1963).

While several algorithms for rigid-body refinement in reciprocal space have been described (Scheringer, 1963; Sussman, Holbrook, Church & Kim, 1977; Huber & Schneider, 1985), the procedure described here operates in real space. The error function that is minimized is the sum of the squared differences between the observed and calculated structure-factor amplitudes. Derivatives of the error function with respect to atomic positions are obtained from the *DERIV* programs of Jack & Levitt (1978). The difference Fourier map required by the *DERIV* programs is calculated with a fast Fourier transform (FFT) (Cooley & Tukey, 1965; Ten Eyck, 1973). In addition, the program described here (*MATREF*) differs fundamentally from other approaches in the way the geometric transformation (rotation and translation) is represented. Previous methods have represented the transformation in terms of Eulerian angles, which may be inconvenient and difficult to constrain for a specific problem. In contrast, the present treatment considers only the linear matrix representation of the transformation. This formulation makes the algorithm extremely versatile, and eliminates the need to perform any trigonometric calculations.

2. Matrix element refinement

Twelve parameters, p, are required to specify a general transformation; nine for the 3×3 linear transformation matrix and three for the translation. The twelve parameters take the coordinate vector (x_1^i, x_2^i, x_3^i) into $(x_1^{i+1}, x_2^{i+1}, x_3^{i+1})$ through the relationship

$$\begin{bmatrix} x_1^{i+1} \\ x_2^{i+1} \\ x_3^{i+1} \end{bmatrix} = \begin{bmatrix} p_1 & p_4 & p_7 & p_{10} \\ p_2 & p_5 & p_8 & p_{11} \\ p_3 & p_6 & p_9 & p_{12} \end{bmatrix} \begin{bmatrix} x_1^i \\ x_2^i \\ x_3^i \\ 1 \end{bmatrix} = P \begin{bmatrix} x_1^i \\ x_2^i \\ x_3^i \\ 1 \end{bmatrix} \quad (1)$$

where the superscript, i or $(i+1)$, denotes the cycle of refinement. The $p_1, p_2, \ldots, p_{12}$ represent the incremental transformation for one cycle. If desired, the linear transformation matrix specified by the first nine parameters of P may be factored into a rotational component and a triangular strain component, which may be related to errors in unit-cell parameters.

3. Constraints and type of refinement

The parameters p_i are refined under linear user-defined relationships between the parameters. This introduces tremendous flexibility in the algorithm. Refinements as constrained as pure translation or as free as rigid-body and unit-cell refinement can be specified easily. Assuming that the desired refinement has n degrees of freedom, the user specifies a $12 \times n$ constraint matrix, C, that relates the twelve param-

*Present address: Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037, USA.

0021-8898/88/060925-04$03.00

eters to the n free variables $(f_1, f_2, \ldots, f_n)$ by

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_{12} \end{bmatrix} = C \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} + \mathbf{i}. \tag{2}$$

The vector $\mathbf{p}$ is the vector representation of P and $\mathbf{i}$ is the vector representation of the identity transformation,

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3}$$

so that when all $f_k = 0$, $P_{ij} = \delta_{ij}$.

The first and second derivatives of the error function with respect to the atomic positions are obtained from *DERIV*.* These derivatives and the linear relationships in (1) and (2) are used by *MATREF* to compute the first and second derivatives of the error function with respect to the free variables, f_k. The resulting gradient–curvature equation,

$$\begin{bmatrix} \frac{\partial^2 E}{\partial f_1^2} & \frac{\partial^2 E}{\partial f_1 \partial f_2} & \cdots & \frac{\partial^2 E}{\partial f_1 \partial f_n} \\ \frac{\partial^2 E}{\partial f_2 \partial f_1} & \frac{\partial^2 E}{\partial f_2^2} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 E}{\partial f_n \partial f_1} & \cdots & & \frac{\partial^2 E}{\partial f_n^2} \end{bmatrix} \mathbf{f} = - \begin{bmatrix} \frac{\partial E}{\partial f_1} \\ \frac{\partial E}{\partial f_2} \\ \vdots \\ \frac{\partial E}{\partial f_n} \end{bmatrix} \tag{4}$$

is solved to give $f_1, f_2, \ldots, f_n$. The twelve parameters, p, are then given by (2). This transformation is then compounded with the current transformation (accumulated in previous cycles) to give the new cumulative transformation.

The errors or uncertainties in the parameters may be estimated, and are reported by *MATREF*. Refinement has converged when the shifts in the parameters are similar in magnitude to the errors. The final errors represent a combination of experimental uncertainty and thermal motion of the molecules. An iterative scheme for refining the orientational parameters is outlined in Fig. 1. The following sections demonstrate the versatility of the algorithm by describing examples of three different types of refinement.

3.1. *Rotation and translation*

For this case, there are six degrees of freedom; three for rotation and three for translation.* Since the difference between an infinitesimal rotation matrix and the identity matrix is antisymmetric (Courant & Hilbert, 1953), the p must be of the form

$$P = \begin{bmatrix} 1 & -f_1 & -f_2 & f_4 \\ f_1 & 1 & -f_3 & f_5 \\ f_2 & f_3 & 1 & f_6 \end{bmatrix}. \tag{5}$$

The constraint matrix, C, becomes

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

The assignment of the f_i in (5) is partially arbitrary, with different assignments corresponding to interchange of the columns of C. The rows of C may not be interchanged.

The linear approximation for the rotation matrix is only valid in the infinitesimal limit. Consequently, if one wishes to refine a true rotation, the rotational component should be factored from the transformation matrix. In this way, the non-linearity of the problem is treated, external to the derivative calculations, by a single matrix factorization. In contrast, methods that deal with Eulerian angles involve extensive trigonometric calculations.

3.2. *Dimer on a crystallographic twofold axis*

Suppose a dimeric structure is known to sit in a monoclinic unit cell (c unique) with its molecular symmetry axis coincident with the crystallographic symmetry axis. There is only one degree of freedom,

*The *DERIV* programs only provide a block-diagonal approximation (with independent atoms) of the second derivatives with respect to atomic positions. The elements of the second derivative (or Hessian) matrix in (4) are computed from these values, and are therefore approximate.

*Even in space groups where some of the translational elements are irrelevant, the block-diagonal nature of the approximation used to obtain the Hessian matrix of (4) makes it necessary to include the translational elements as degrees of freedom. Although the Hessian matrix is never degenerate in practice, eigenvalue filtering is still employed as a precautionary measure.

which corresponds to rotation of the dimer about the z axis. Infinitesimal rotation about the z axis is represented by non-zero elements p_2 and p_4. P takes the form

$$P=\begin{bmatrix} 1 & -f_1 & 0 & 0 \\ f_1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (7)$$

and the matrix C takes the form

$$C=\begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (8)$$

As before, the true rotational component of the resulting transformation may be extracted by matrix factorization.

3.3. *Rotation, translation, and variable orthorhombic unit cell*

If a molecular-replacement model is placed in a unit cell with slightly incorrect cell parameters, it will be impossible for atoms in the model to reach their true fractional coordinates by rigid-body motion. We may overcome this problem by temporarily allowing the transformation to stretch the model slightly in order for the atoms to reach more correct fractional positions in the unit cell. The transformation may be factored into a rotational component and a triangular component by Gram–Schmidt orthogonalization (Strang, 1980). If the strain in the model is a result of unit-cell error or variability, then the triangular component should obey the lattice symmetry. This condition may be imposed on the infinitesimal transformation, as illustrated for a model in an orthorhombic unit cell. The infinitesimal transformation should be a product of an infinitesimal rotation and an infinitesimal diagonal strain. P takes the form

$$P=\begin{bmatrix} 1+f_7 & -f_1 & -f_2 & f_4 \\ f_1 & 1+f_8 & -f_3 & f_5 \\ f_2 & f_3 & 1+f_9 & f_6 \end{bmatrix}. \quad (9)$$

The constraint matrix C becomes

$$C=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \quad (10)$$

If significant strain is detected, and the original model is believed to be free of strain, then it is likely that the unit cell in which the model is being refined is slightly erroneous. Only the rotational and translational components of the transformation should be applied to the model.

The strain component of the transformation can be used to adjust the unit cell, since the derivatives of structure factors with respect to unit-cell parameters are very nearly proportional to derivatives with respect to elements of the triangular component of the transformation matrix. Updated unit-cell parameters are obtained by first generating the matrix for converting Cartesian coordinates to fractional coordinates, then multiplying by the triangular strain matrix, and extracting new unit-cell parameters from the resulting matrix.

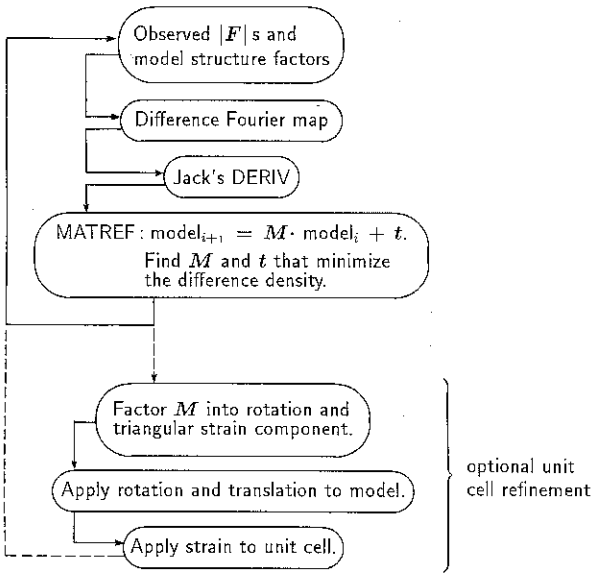

Fig. 1. Flowchart for real-space rigid-body refinement. Refinement of unit-cell parameters is optional. M and $\mathbf{t}$ are the matrix and translational components of the transformation P in the text.

4. Results and discussion

The algorithm described here has been used successfully in the rigid-body refinement of a bacterial photo-

synthetic reaction center (Allen, Feher, Yeates, Komiya & Rees, 1987). For the case of the reaction center, variability in the unit-cell parameters made it desirable to optimize these parameters, as well as the orientation of the model. Four cycles of rigid-body refinement resulted in a rotation of 1·9° and a shift of 1·4 Å in the center of mass of the model. The diagonal component of the strain matrix was factored from the transformation and used to adjust the unit-cell parameters in each cycle, as diagrammed in Fig. 1. The cumulative effect was to change the orthorhombic unit cell from 142·2, 75·6, 141·8 Å to 138·0, 77·5, 141·8 Å. The change in the unit cell was significant. The R factor between the observed data and the structure factors calculated from the original model was 50% for data between 8 and 3·5 Å. The R factor for the new model (rotated, translated and placed in the new unit cell) was 43%.

Group refinement is performed by first calculating the derivatives of the error function with respect to all the atomic positions, as in rigid-body refinement. These derivatives are then used to refine an independent transformation for each group of atoms. Most of the computation time involves calculation of the derivatives, with very little time required for *MATREF* to calculate the new transformations. Consequently, the time required to perform group refinement is fairly insensitive to the number of independent groups being refined.

The programs have been used in the group refinement of an Fab fragment of an antipeptide antibody (Rini & Wilson, 1988). Rigid-body refinement was followed by group refinement with two, four and finally eight domains. The initial R factor was 63% for data between 12 and 8 Å resolution and the final R factor was 42% for data between 10 and 4 Å. For comparison, the program *CORELS* (Sussman *et al.*, 1977) was used to carry out the same refinement, with very similar results. The two methods converged in approximately the same number of cycles. With 4290 atoms in four domains, and data between 10 and 5 Å, *CORELS* required 390 min per cycle on a VAX11/750, while the real-space procedure described here required 31 min.

5. Concluding remarks

The procedure presented here is noteworthy because of the speed and versatility with which it can perform rigid-body and group refinement. The procedure owes its speed to the method used to calculate the derivatives of the error function (summation of the squared differences between calculated and observed structure-factor amplitudes) with respect to atomic positions. The derivatives are obtained in real space by evaluating the gradient in a difference Fourier map (calculated with a FFT) at the atomic positions of the model, using the *DERIV* programs of Jack & Levitt (1978). The program *MATREF* then refines the orientation of the model on the basis of these derivatives.

While FFTs may also be employed in reciprocal-space rigid-body refinement (Huber & Schneider, 1985), such a procedure is not easily extended to group refinement. A separate Fourier transform would be required for each group, and the computation time would be nearly proportional to the number of groups. As previously described, the real-space method allows multiple groups to be refined using a single FFT in each refinement cycle.

The versatility of the method results from the matrix representation of the transformation of the model or group. The derivatives of the error function with respect to the matrix elements are calculated from the derivatives with respect to the atomic positions. In addition to avoiding the difficulties associated with Eulerian angles, the linear representation allows the user to specify many different types of refinement by the use of a small constraint file. The matrix representation also allows the user to refine the values of unit-cell parameters by allowing the transformation to contain a component of strain, which is then factored from the transformation and used to adjust the unit-cell parameters as previously described.

We wish to thank J. M. Rini and M. Lewis for helpful discussions and the referees for useful comments. This work was supported, in part, by USPHS grant GM31299. *MATREF* and additional useful programs for real-space rigid-body refinement are available upon request.

References

ALLEN, J. P., FEHER, G., YEATES, T. O., KOMIYA, H. & REES, D. C. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 5730–5734.
COOLEY, J. W. & TUKEY, J. W. (1965). *Math. Comput.* **19**, 297–301.
COURANT, R. & HILBERT, D. (1953). *Methods of Mathematical Physics*, Vol. 1, pp. 41–42. New York: Wiley.
HUBER, R. & SCHNEIDER, M. (1985). *J. Appl. Cryst.* **18**, 165–169.
JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931–935.
RINI, J. M. & WILSON, I. A. (1988). Unpublished results.
SCHERINGER, C. (1963). *Acta Cryst.* **16**, 546–550.
STRANG, G. (1980). *Linear Algebra and its Applications*, 2nd ed., pp. 128–131. New York: Academic Press.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. (1977). *Acta Cryst.* A**33**, 800–804.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.